

## **General Disclaimer**

### **One or more of the Following Statements may affect this Document**

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.
- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.
- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.
- This document is paginated as submitted by the original source.
- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Available under NASA sponsorship  
Interest of early and wide dis-  
tribution of Earth Resources Survey  
data for information and without restriction  
to any use made thereof

7.9-10100

CR-158056

(E79-10100) CLASSIFICATION ON ACCURACY OF  
DIFFERENT OPTICS OF THE IMAGE-100 SYSTEM  
(Instituto de Pesquisas Espaciais, Sao Jose)  
17 p HC A02/MF A01 CSCL 14E

N79-18374

Unclass

G3/43 00100



CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO

INSTITUTO DE PESQUISAS ESPACIAIS

"Made available under NASA sponsorship  
in the interest of early and wide dis-  
semination of Earth Resources Survey  
Program information and without liability  
for any use made thereof."

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                           |                                                                                |                                                                                       |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------|--------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| 1. Classification INPE-COM.10/PE<br>C.D.U.: 621.38 SR: 711.14                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                           | 2. Period                                                                      | 4. Distribution<br>Criterion                                                          |
| 3. Key Words (selected by the author)<br><br>LAND USE<br>CLASSIFICATION ACCURACY<br>REMOTE SENSING                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                           |                                                                                | internal <input type="checkbox"/><br><br>external <input checked="" type="checkbox"/> |
| 5. Report No.<br>INPE-1390-PE/182                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | 6. Date<br>November, 1978 | 7. Revised by<br><i>E. R. Souza</i><br>Celso de Renna e Souza                  |                                                                                       |
| 8. Title and Sub-title<br><br>CLASSIFICATION ACCURACY OF DIFFERENT OPTIONS<br>OF THE IMAGE-100 SYSTEM                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                           | 9. Authorized by<br><br><i>N. Parada</i><br>Nelson de Jesus Parada<br>Director |                                                                                       |
| 10. Sector DSE/DIN                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | Code                      | 11. No. of Copies 15                                                           |                                                                                       |
| 12. Authorship R. Kumar<br>M. Niero                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                           | 14. No. of Pages 16                                                            |                                                                                       |
| 13. Signature of first author <i>R. Kumar</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                           | 15. Price                                                                      |                                                                                       |
| 16. Summary/Notes<br>The purpose of this study was to compare the classification accuracy of land use classes of São José dos Campos, SP, Brazil, using the different options of signature acquisition for classification available in the Image-100, a system developed by General Electric Co., pixel-by-pixel maximum likelihood gaussian classifier (MLGC), and the sample classifier. In addition, the statistical separability of land use classes in the subsets of one to four spectral channels was investigated. With the help of ground observations and aerial photography, the multispectral scanner (MSS) data of LANDSAT were analysed using the Image-100. For the single-cell option of the Image-100, the errors of omission varied from 16.3% for the class "commercial" to 26.8% for the class "residential". The errors of commission varied from 5.6% for the class "commercial" to 33.2% for the class "Unoccupied". As expected, the multi-cell option increased the errors of omission and decreased the errors of commission. On the whole, the sample classifier gave slightly more accurate results than MLGC and much more accurate than any of the options of classification available in Image-100. |                           |                                                                                |                                                                                       |
| 17. Remarks To be submitted for publication in the IEEE Transactions on Geoscience Electronics.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                           |                                                                                |                                                                                       |

## INDEX

|                          | Page      |
|--------------------------|-----------|
| LIST OF TABLES .....     | <i>iv</i> |
| ABSTRACT .....           | 1         |
| INTRODUCTION .....       | 2         |
| LITERATURE REVIEW .....  | 3         |
| METHOD OF ANALYSIS ..... | 3         |
| REFERENCES .....         | 10        |

PRECEDING PAGE BLANK NOT COUNTED

## LIST OF TABLES

|                                                                       | Page |
|-----------------------------------------------------------------------|------|
| Table I - Values of $B_{AVE}$ in subset of one to four channels ..... | 12   |
| Table II - Percentage errors of omission and commission .....         | 13   |

# CLASSIFICATION ACCURACY OF DIFFERENT

## OPTIONS OF THE IMAGE-100 SYSTEM

R. Kumar and M. Niero\*

### ABSTRACT

The purpose of this study was to compare the classification accuracy of land use classes of São José dos Campos, SP, Brazil, using the different options of signature acquisition for classification available in the Image-100, a system developed by General Electric Co., pixel-by-pixel maximum likelihood gaussian classifier (MLGC), and the sample classifier. In addition, the statistical separability of land use classes in the subsets of one to four spectral channels was investigated. With the help of ground observations and aerial photography, the multispectral scanner (MSS) data of LANDSAT were analysed using the Image-100. For the single-cell option of the Image-100, the errors of omission varied from 16.3% for the class "commercial" to 26.8% for the class "residential". The errors of commission varied from 5.6% for the class "commercial" to 33.2% for the class "unoccupied". As expected, the multi-cell option increased the errors of omission and decreased the errors of commission. However, considering both the errors of omission and commission, this option considerably decreased the percentage of correct classification as compared to the single-cell option. On the whole, the sample classifier gave slightly more accurate results than MLGC and much more accurate than any of the options of classification available in Image-100.

---

\* The authors are with the Instituto de Pesquisas Espaciais (INPE), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), 12200 - São José dos Campos, SP, Brazil.

## INTRODUCTION

The purpose of this study was to compare the classification accuracy of land use classes using the different options of signature acquisition for classification available in the Image-100 (Image-100 is a data processing system marketed by General Electric Co. that extracts thematic information from multispectral imagery, enhances the image, etc.). In addition, the statistical separability of land use classes in the subsets of one to four spectral channels was investigated.

Cloud free multispectral scanner data from LANDSAT, of reasonable quality over São José dos Campos ( $23^{\circ} 10' S$ ,  $45^{\circ} 50' W$ ), São Paulo, Brasil, acquired on September 8, 1972, were available. In addition, aerial photography and ground observations were available to assist the analysis of the data. São José dos Campos was selected because it is one of the fastest growing small-size towns of Brasil and the authors are well familiar with it. Many of the problems of this town are similar to the problems of much larger urban centers.

With the help of ground observations and aerial photography, a map of São José dos Campos, showing the following land use classes was obtained: residential areas, commercial areas, agricultural areas and unoccupied areas.

The specific objectives of the study are stated as follows:

1. To determine what combinations of one through three spectral channels out of four available channels give the greatest overall statistical separability of the above four land use classes.
2. To compare the classification accuracy of land use classes using 'single-cell signature acquisition' and 'multi-cell signature acquisition' options of classification available in

the Image-100, pixel-by-pixel maximum likelihood gaussian classifier (MLGC) and a sample classifier, on-line-mode, in the Image-100.

## LITERATURE REVIEW

Many investigators have analysed the multispectral scanner (MSS) data of LANDSAT satellite for applications to land use classification. For example, Todd and Baumgardner<sup>1</sup> (1973) analysed LANDSAT MSS data obtained over Marion County (Indianapolis), Indiana, by computer-implemented techniques to evaluate the utility of satellite data for urban land use classification. Several land use classes, such as commerce/industry, single-family (newer) residential, trees, and water exhibited spectrally separable characteristics and were identified with greater than 90 percent accuracy. Ellefsen et al.<sup>2</sup> (1973) did computer-aided analysis of LANDSAT MSS data of the San Francisco Bay area. Smith et al.<sup>3</sup> (1974) have given the application of spatial features to satellite land-use analysis. Ellefson et al.<sup>4</sup> (1974) have given new techniques in mapping urban land use and monitoring change for selected U.S. metropolitan areas. They analysed LANDSAT MSS data using automatic pattern recognition techniques for classification. Kumar and Silva<sup>5</sup> (1977) have analysed the statistical separability of agricultural cover types in much detail, data quantity and depth in the subsets of one to twelve spectral channels.

Goldberg et al.<sup>6</sup> (1975) have described methods and procedures which outside investigators may use with the automated processing equipment of the Canada Centre for Remote Sensing (CCRS) for the purpose of natural resource exploration and mapping. They have compared the accuracies of unsupervised and supervised methods on the basis of the confusion matrices generated by classifying exactly the same area.

## METHOD OF ANALYSIS

Multispectral scanner data of computer compatible tapes of LANDSAT were analysed using Image-100. With the aid of land use map



of São José dos Campos, mentioned above, rectangular areas of each of the above four land use classes were selected, avoiding the boundaries of classes on the display of the Image-100. The areas of each of these classes were selected carefully so that they could be considered to be representative of the respective land use classes. Assuming that each of these classes has a multivariate gaussian distribution, the B-distance based on Bhattacharyya coefficient was calculated between all possible pairs of these classes in all possible combinations of one, two, three and four spectral channels using the feature selection algorithm of the Brazilian Institute of Space Research (INPE), on-line-mode, with the Image-100<sup>7,8</sup>. For each value of B-distance, the probability of correct classification was reasonably estimated from the curve of Swain and King (1973)<sup>9</sup>. The B-distance for two multivariate gaussian distribution is given by<sup>9</sup>:

$$B = 2 (1 - e^{-\alpha}), \quad (1)$$

where

$$\alpha = \frac{1}{8} (U_1 - U_2)^T \Sigma^{-1} (U_1 - U_2) + \frac{1}{2} \log_e \left[ \frac{\det \Sigma}{\sqrt{\det \Sigma_1 \cdot \det \Sigma_2}} \right] \quad (2)$$

where  $U_1$  and  $U_2$  are mean vectors of classes one and two respectively; whereas,  $\Sigma_1$  and  $\Sigma_2$  are the covariance matrices of the same two classes.

$$\Sigma = \frac{1}{2} [\Sigma_1 + \Sigma_2] \text{ and } T \text{ denotes transpose} \quad (3)$$

The average B-distance over all pairs of classes is given by:

$$B_{AVE}(C_1, C_2, \dots, C_n) = \frac{2}{m(m-1)} \sum_{i=1}^{m-1} \sum_{j=i+1}^m B(i, j | C_1, C_2, \dots, C_n) \quad (4)$$

where

$m$  = number of classes

$B(i, j | C_1, C_2, \dots, C_n)$  = B-distance between classes  $i$  and  $j$  in the channels  $C_1, C_2, \dots, C_n$ .

$B_{AVE}$  was computed for all possible subsets of one, two, three and four spectral channels out of the available four channels.

Each of these land use classes was divided into two independent sets: training fields and test fields. Using training fields of residential areas, test fields of each of the above four classes were classified using the single-cell signature acquisition option of Image-100. This option creates a four-dimensional rectangular parallelepiped, each side of which corresponds to the signature limits of the training areas in each channel. The number of pixels classified as residential areas by the computer inside the test fields of each of these four classes were determined. An identical analysis was repeated for each of the other three land use classes. Thus, a confusion matrix showing the total number of pixels (picture elements) of each class classified correctly as well as classified incorrectly into each of the other classes was obtained.

This whole procedure was repeated for the multicell signature acquisition option of the Image-100. In the multicell signature acquisition, the parallelepiped of spectral signature is subdivided into cells, each of unit volume, and the number of pixels in each of these unit cells is counted. These cell counts are, thus, measures of the probability distribution of the spectral cluster. By raising or lowering the threshold on the cell counts, one can vary the size of the four dimensional probability distribution of the spectral cluster by deleting or adding cells with counts greater than the variable threshold. In the interactive signature modification option, the user

performs training on the misclassified area, adding the errors of omission and subtracting the errors of commission until satisfied with the results.

The same training fields of each class were used to classify the test fields using MLGC as well as the sample classifier based on B-distance. B-distance was computed between a test field and each of the four training classes, and the field was classified into the class for which the B-distance was minimum.

Table 1 gives the values of  $B_{AVE}$  in all possible combinations of one, two, three and four channels out of the four available channels. As one would expect, the values of  $B_{AVE}$  increase with an increase in the number of channels. In the subsets of one to three spectral channels, channel 4, channels 4 & 7 (one in the visible and one in the near infrared), and channels 4, 5 & 7 (two in the visible and one in the near infrared) are found to be the best choices. Table 1 shows that in the subset of two channels, channels 4 and 5 (visible wavelength region) give higher probability of correct classification than channels 6 & 7 (near infrared wavelength region). The authors believe that each wavelength region -- visible, near infrared, middle infrared and thermal infrared has independent information content. Thus, in the subset of two spectral channels, one channel in the visible and one channel in the near infrared wavelength region are found to be the best choice. Kumar (1978)<sup>10</sup> has analysed aircraft collected MSS data in much detail, data quantity and depth in the subsets of one to twelve spectral channels to evaluate each spectral channel as well as possible combinations of wavelength regions for statistical separability of agricultural cover types.

The errors of omission (while using training fields of residential areas, number of pixels of test fields known to be residential, not classified as residential constitute the errors of omission, etc.) and the errors of commission (while using training fields of residential areas, number of pixels of classes other than residential but which are classified by the Image-100 as residential) were calculated and are shown in Table 2. Similarly, the errors of

omission and commission using the multicell signature acquisition ( $m=1$ ,  $m=2$  and  $m=3$ ), for the *same* training and test fields of each class were calculated and are given in Table 2. The option  $m = 1$  means that all the unit cells in the four dimensional spectral space, which had less than one pixel, were deleted from the spectral signature of the training fields for doing classification. Similarly, the option  $m = 2$  means that all the unit cells in the four dimensional spectral space which had less than two pixels were deleted from the spectral signature of the training fields for doing classification etc. Table II shows that for the single-cell option, the errors of omission vary from 16.3% for the class commercial to 26.8% for the class residential. The errors of commission vary from 5.6% for the class commercial to 33.2% for the class unoccupied. It shows that classification accuracy for all the classes is rather poor except the class commercial where the percentage of errors are reasonably small (errors of omission = 16.3%, commission = 5.6%). This is because of small values of standard deviation of this class (and hence, less overlap with other classes) in each of the spectral channels, specially in the channels one (0.5 to 0.6  $\mu\text{m}$ ) and four (0.8 to 1.1  $\mu\text{m}$ ).

In general, an increase in the standard deviations of a class in the spectral channels tends to reduce the errors of omission and increase the errors of commission. It was found that taking into account both the errors of omission as well as those of commission, the classification accuracy generally decreases with an increase in the standard deviations.

Table II shows, as expected, that the multicell option increases the errors of omission and decreases the errors of commission. Considering the errors of omission as well as errors of commission, multicell option for  $m = 1$  considerably decreases the percentage of correct classification for each of the classes. This is because the number of pixels used for training of each class were relatively small for statistical purpose. Thus, the unit cells in the four dimensional spectral space were sparsely populated. Thus, there may

be many cells which are actually representative of the class but do not have any pixel because the total number of pixels for training for each of the classes was rather small. For the multicell option, the errors of omission increase and the errors of commission decrease as we go from  $m = 1$  to  $m = 2$  to  $m = 3$ . Considering the errors of omission as well as the errors of commission, the percentage of correct classification decreases as we go from  $m = 1$  to  $m = 2$  to  $m = 3$ .

Table II also shows that the interactive signature acquisition option does not improve the classification accuracy as compared to the "single cell" option because the basic problem is the overlap between the classes in the four-dimensional spectral space.

Table II shows the results of classification using pixel-by-pixel maximum likelihood gaussian classifier (MLGC) as well as a sample classifier. As pointed out earlier, the same training and test fields were used in this case as in the "single cell" or "multicell option" of the Image-100.

Comparing these results to the single cell option, we find that the errors of commission are reduced significantly; whereas, the errors of omission are increased for some classes and decreased for the others. For most classes, the larger were the standard deviations, the lower was the percentage of correct classification using the sample classifier. Comparing the sample classifier to the 'multicell option', we find that it gives much smaller errors of omission. However, MLGC gives higher errors of commission; whereas, the sample classifier gives greater errors of commission for some classes and smaller for the others. On the whole, the sample classifier gives a percentage of correct classification slightly better than MLGC and much better than the single cell or multi-cell options of the Image-100..

In the future, computer compatible tapes being developed at INPE of São José dos Campos of other times will be analysed to investigate the effect of time on these results.

The authors gratefully acknowledge the assistance of Dr. Nelson de Jesus Parada, Director of the Instituto de Pesquisas Espaciais (INPE); Dr. Celso de Renna e Souza, Mr. J.C. Moreira and Mrs. M. S.S. Barros of INPE for their assistance with this work.

REFERENCES

1. W.J.Todd and M.F. Baumgardner, "Land use classification of Marion County, Indiana by spectral analysis of digitized satellite data", Conference on Machine Processing of Remotely Sensed Data, Laboratory for Applications of Remote Sensing, Purdue University, W. Lafayette, Indiana, October 16-18, 1973.
2. R. Ellefsen, P.H. Swain, J.R. Wray, "Urban land use mapping by machine processing of ERTS-1 multispectral data: A San Francisco Bay Area example", Conference on Machine Processing of Remotely Sensed Data, Laboratory for Applications of Remote Sensing, Purdue University, W. Lafayette, Indiana, Oct. 16-18, 1973.
3. J.A. Smith, R.J. Hornung, and J.K. Berry, "Application of spatial features to satellite land-use analysis", Proc.: Sym. on Remote Sensing and Photo Interpretation, the Canadian Institute of Surveying, Banff, Alberta, Canada, pp.23-38, 1974.
4. R. Ellefson, L. Gaydos, P.H. Swain, and J.R. Wray, "New techniques in mapping urban land use and monitoring change for selected U.S. metropolitan areas: an experiment employing computer - assisted analysis of ERTS-1 MSS data", Proc.: Symp. on Remote Sensing and Photo interpretation, The Canadian Institute of Surveying, Banff, Alberta, Canada, pp. 51-64, 1974.
5. R.Kumar and L. Silva, "Separability of agricultural cover types by remote sensing in the visible and infrared wavelength regions", IEEE Trans. Geosci. Elec., vol. GE-15, pp.42-49, 1977.
6. M. Goldberg, D. Goodenough and S. Shlien, "Classification methods and error estimation for multispectral scanner data", Third Canadian Sym. on Remote Sensing, Edmonton, September 22-24, pp. 125-143, 1975.
7. R. Kumar, "Feature selection algorithm", in Collection of relevent results obtained with the Skylab Images by the Institute for Space

Research, INPE-793-NTE/043, São José dos Campos, SP, Brazil, 1975.

8. R. Kumar, "Feature selection and sample classification algorithms of INPE", Proceedings: Sixth Annual Remote Sensing of Earth Resources Conference, The University of Tennessee Space Institute, Tullahoma, Tennessee, 1977.
9. P.H. Swain and R.C. King, "Two effective feature selection criteria for multispectral remote sensing". LARS Information Note 042673, Laboratory for Applications of Remote Sensing, Purdue University, W. Lafayette, Indiana, 1973.
10. R. Kumar, "Evaluation of spectral channels and wavelength regions for separability of agricultural cover types", Proc. Int. Sym. on Remote Sensing of Environ., Environmental Research Institute of Michigan, Ann Arbor, MI, pp. 1081-1090, 1977.



Table I - Values of  $B_{AVE}$  in subset of one to four channels

| Channel | $P_C$ | Channels | $P_C$ | Channels | $P_C$ |
|---------|-------|----------|-------|----------|-------|
| 4       | 84.3  | 4-5      | 85.0  | 4-5-6    | 86.6  |
| 5       | 84.0  | 4-6      | 85.0  | 4-5-7    | 88.5  |
| 6       | 74.5  | 4-7      | 86.1  | 4-6-7    | 86.7  |
| 7       | 74.4  | 5-6      | 85.1  | 5-6-7    | 84.6  |
|         |       | 5-7      | 86.0  | 4-5-6-7  | 89.0  |
|         |       | 6-7      | 79.8  |          |       |

Note:  $P_C$  denotes probability of correct classification estimated from the values of  $B_{AVE}$  using curve of Swain and King<sup>9</sup>.

Table II - Percentage errors of omission and commission

| Class                        | (A) Percentage errors of omission (%) |                     |                     |                     |             | Sample Classifier |
|------------------------------|---------------------------------------|---------------------|---------------------|---------------------|-------------|-------------------|
|                              | Single Cell                           | Multi-cell<br>m = 1 | Multi-cell<br>m = 2 | Multi-cell<br>m = 3 | Interactive |                   |
| Residential                  | 26.8                                  | 63.2                | 73.9                | 83.6                | 23.2        | 4.35              |
| Commercial                   | 16.3                                  | 74.7                | 78.5                | 83.4                | 25.6        | 0                 |
| Agricultural                 | 19.4                                  | 73.5                | 80.3                | 86.7                | 21.1        | 26.8              |
| Unoccupied                   | 16.9                                  | 77.5                | 82.3                | 88.4                | 34.2        | 32.5              |
| (B) Errors of commission (%) |                                       |                     |                     |                     |             |                   |
| Residential                  | 17.7                                  | 2.3                 | 1.6                 | 0.6                 | 15.5        | 1.9               |
| Commercial                   | 5.6                                   | 1.9                 | 1.0                 | 0.4                 | 29.6        | 1.5               |
| Agricultural                 | 30.0                                  | 4.9                 | 2.7                 | 2.1                 | 24.9        | 12.4              |
| Unoccupied                   | 33.2                                  | 7.0                 | 5.3                 | 4.7                 | 38.6        | 11.0              |